TEACHER BIASES IN THE IDENTIFICATION OF LEARNING DISABILITIES: AN APPLICATION OF THE LOGISTIC MULTILEVEL MODEL

Georgios D. Sideridis, Faye Antoniou, and Susana Padeliadu

Abstract. The purpose of the present study was to investigate the presence of teacher biases with regard to identification of students with learning disabilities (LD). Factors related to teachers’ gender, age, and experience, along with children’s gender, were investigated. Results suggested that teachers’ gender is associated with biases with regard to identification of learning disabilities by a factor of 2:1. In other words, every child who is rated by female teachers as having an LD (who actually has LD) corresponds two children when rated by male teachers. Students’ gender, on the other hand, did not differentially predict identification rates. Furthermore, teacher age and experience did not contribute significantly to student identification rates. The findings are discussed with regard to policy mandates and classification schemes.

GEORGIOS D. SIDERIDIS, Ph.D., Department of Psychology, University of Crete.
FAYE ANTONIOU, Ph.D., Department of Special Education, University of Thessaly.
SUSANA PADELIADU, Ph.D., Department of Special Education, University of Thessaly.

Teachers’ reports are often used to assess various behaviors in students with and without disabilities (Liljequist & Renk, 2007; Wingenfeld, Lackar, Gruber, & Kline, 1998). Based on the premise that teachers are very important adults in students’ lives (Achenbach, 1991), they are asked to contribute their unique perspective regarding various aspects of students’ behavior and achievement from preschool to college (Jensen, 1980; Mashburn & Henry, 2004; Temp, 1971). For example, their ratings are used for identification purposes, classification, or just assessment. Regardless of usage, these ratings carry information regarding teachers’ predictions of how students will behave and achieve in the future (Wood & Benton, 2005).

Although there is evidence that general and special education teachers often “miss” characteristics defining a disability (Fox & Lent, 1996; Miller, Missiuna, Macnab, Malloy-Miller, & Polatajko, 2001), the importance of their ratings is not debated (Rosenthal & Jacobson, 1968; Rubovits & Maehr, 1973; Tenenbaum & Ruck, 2007), particularly because their estimations are core predictors of students’ placement (Podell & Soodak, 1993; Proctor & Prevatt, 2003).

In the present study, we evaluated the role of teachers in identifying students with learning disabilities (LD) and examined the possible presence of bias on their part, including the moderating role of their personal characteristics. The term bias in this study refers to the systematic difference in the ways by which teachers identify learning disabilities and the factors related to them. Those ways may affect teachers’ assessment of student performance.
Teachers as a Means of Identifying Students with Disabilities

Teacher ratings are often used for identification of students suspected of having various disabilities. Teachers’ involvement in the identification process of students with LD may now be greater than in the past, given the use of the responsiveness-to-intervention (RTI) model of identification (Haager, Calhoun, & Linan-Thompson, 2007; Kavale, Holdnack, & Mostert, 2006; Vaughn & Fuchs, 2003). As Kavale et al. (2006) pointed out, the decision regarding identification is based on “‘clinical’ (i.e., teacher) judgements about the level of response” (p. 120).

Judgments on the part of teachers are based on their expectations, perceptions, skills, and training. It is imperative for teachers to provide unbiased estimates of their students’ ability, potential, and achievement. Several researchers have blamed teachers for overidentification of learning disabilities (e.g., Wong, 1996). Others have suggested that the number of students with mental retardation has declined primarily because of overidentification of learning disabilities (MacMillan, Siperstein, & Gresham, 1996) (see also Keogh, 2005).

Nevertheless, teacher ratings commonly assist in identifying learning disabilities (e.g., The Learning Disability Diagnostic Inventory [LDDI]; Hammill, 1995) and attention deficit hyperactivity disorder (e.g., Dupaul et al., 1998). A crucial question is how well teachers accomplish their task, given the often discrepant findings between ratings and observational protocols (Lorenz, Melby, Conger, & Xu, 2007). For example, investigating the interrelationships between intelligence, academic achievement, and teacher estimations of that achievement, Svanum and Bringle (1982) found that teacher race or socioeconomic status did not affect the predictability of their ratings. This finding was replicated by Abidin and Robinson (2002).

In contrast, Messe, Crano, Messe, and Rice (1979) reported that teacher ratings are biased by students’ socioeconomic status. Social class and race seemed to be the main factors that influence these ratings (Knotek, 2003; McLeskey, Waldron, & Wornhoff, 1990; Wehmeier & Schwartz, 2001) as well as placement (general vs. special education). That is, teachers in general education settings appear to be less reliable when providing ratings related to LD than special education teachers in special education settings (Vaughn & Fuchs, 2003).

Chang and Demyan (2007) also studied the presence of minority stereotypes in teacher ratings. The authors noted, “Contrary to expectation, there was substantial congruence in the degree of uniformity and favorableness of the stereotypic traits associated with blacks and whites, with participants revealing both strong positive and strong negative trait associations” (p. 91). The authors suggested the need for alternative methods of assessment that may produce more valid ratings as stereotypical ratings can be catastrophic for students’ educational experience. For example, teacher biases have been evident with minority groups (e.g., not calling on Asian-American students as often as they did Caucasian students; see Schneider & Lee, 1990).

Rescorla et al. (2007) examined the consistency of teacher-reported problems for students in 21 countries. Using a sample of 30,957 teachers, results indicated that participating educators were fairly accurate in identifying students with behavioral and emotional problems. Internal consistency estimates averaged .90, and gender effects (i.e., discrepancies or biases) ranged between 1% and 5%, which were pretty small and within the range of what constitutes reliable responding. Overall, the authors concluded that teachers’ reports were similar and consistent across countries, supporting a model of teachers as valid raters of student behavior.

Another noteworthy finding is that parents’ unwillingness to respond to questionnaires is associated with higher bias on the part of teachers regarding the presence of social, emotional, attentional problems, and so forth, in their children (Stormark, Heiervang, Heimann, Lundervold, & Gillberg, 2008). In other words, this pattern of behavior from parents was linked to biased estimates on the part of teachers.

Thus, research findings regarding the effectiveness of teachers in rating students’ attributes and behaviors have been largely inconclusive, leaving unanswered such questions as: Which factors influence teachers’ ratings? Are there specific teacher characteristics that play a prominent role in the validity of their ratings?

Teachers’ Characteristics as Predictors of Their Ratings

Studies have shown that teachers’ characteristics influence the validity of their ratings. For example, Rivard, Missiuna, Hanna, and Wishart (2007) investigated teachers’ accuracy in identifying children with coordination and behavioral problems. Using a sample of 152 Canadian teachers and eight scenarios demonstrating different levels of motor and behavior problems, findings revealed differential effects due to teacher gender. Specifically, female teachers were more likely to underscore the severity of gross-motor problems; the opposite was true of male teachers. With regard to fine-motor skills, males provided lower responses (were less concerned with) than female teachers, underscoring the significant interaction between teacher gender and type of motor problem with regard to degree of concern.

An interaction between teacher gender and inappropriate student behavior was noted by Taylor, Gunter,
and Slate (2001) using videotaped scenarios showing students acting out. They engaged 186 teachers to view the scenarios and to complete a rating scale on inappropriate behaviors. Teacher gender was once again the most influential factor on perceptions. Specifically, female teachers more often referred students with learning and behavioral problems; this was especially true for boys (Hopf & Hatzichristou, 1999). However, Hopf and Hatzichristou found that female educators were more sensitive than their male counterparts. This effect could be attributed to the fact that female teachers may be more expressive when interacting with students (Meece, 1987), although they pay more attention to male than to female students (Bellamy, 1994). Duffy, Warren, and Walsh (2001) reported that, regardless of the subject matter taught, female teachers interacted heavily with male students compared to female students. Regarding behavioral assessment, Ritter (1989) documented that male teachers were less accurate in identifying students’ behavioral problems than female teachers, who showed more sensitivity in distinguishing problem behaviors.

In terms of teaching experience, Stevens, Quittner, and Abikoff (1998) reported that special education teachers rated behavioral problems as less of a burden than did general education teachers. Further, according to Mashburn and Henry (2004), the higher the level of the teachers’ education, the more accurate their ratings. In their study, preschool and kindergarten teachers rated children’s skills to evaluate their readiness to attend school. Teachers with low levels of education and training produced systematically inflated estimations and invalid ratings, and preschool teachers’ ratings were less valid than kindergarten teachers’ estimates. Previous studies also noted that more experienced elementary and secondary teachers are more able to handle students’ challenging behaviors (Borg & Falzon, 1998; Kokkinos, Panayiotou, & Davazoglou, 2005).

The most common characteristic that teachers rely upon in making their judgments seems to be their experience (Brennan O’Neil & Liljequist, 2002). Thus, teacher experience, as reflected by years of teaching in general education classrooms and age, comprised variables that were examined as predictors of classification rate. Neither knowledge of the disorders nor professional experience and educational background seems to play an important role in teachers’ accuracy of attention to deficit symptoms or oppositional defiant disorder (Stevens et al., 1998).

There is also compelling evidence to suggest that teachers’ ethnicity does not influence their judgments (Abidin & Robinson, 2002). Judging the correctness of the referral estimations of almost 200 teachers with diverse ethnic backgrounds, Tobias and his colleagues (Tobias, Cole, Zibrin, & Bodlakova, 1982) found that neither teachers’ nor students’ races were significant predictors of biased estimations of student ability. This was not the case, however, in another study by the same researchers (Tobias, Zibrin, & Menell, 1983) in which they concluded that teachers of a specific race estimated students of their own race as less likely to have special educational needs than students of a different race. In another investigation, Tobias et al. (1983) found that this could be a result of teachers’ experience in special education settings.

Attitudes towards the placement of students with learning disabilities may affect not only teachers’ estimations but also the quality of their interactions with the students (Bender, Vail, & Scott, 1995). The quality of this interaction is not predictive by teacher characteristics, such as perceptions of self-efficacy, knowledge of special education (Bender, Vail, & Scott, 1995; Stevens, Quittner, & Abikoff, 1998), but mostly by students’ gender (Anderson, 1997).

**Teachers Biases with Regard to Student Gender**

There is also evidence suggesting that more boys than girls are identified as having a disability (Anderson, 1997; Shaywitz, Shaywitz, Fletcher, & Escobar, 1990). Regardless of whether the identification refers to neurological (e.g. autism, attention-deficit/hyperactivity, speech and language disorders) or learning and reading disabilities, boys are more often referred for special education services than girls (Lieberman, Kastrowitz, & Flannery, 2005). The disproportionate placement in terms of student gender has been so salient that teacher biases have been suspected (Gillberg, 2003). For example, Rivard et al. (2007) observed differential effects with regard to student gender. Specifically, teachers were more concerned about gross-motor problems in boys than in girls. Conversely, they were more concerned with fine-motor problems in girls than in boys.

Overidentification of boys has been replicated both in younger groups of children and in diverse special education fields. For example, Mashburn and Henry (2004) reported that preschool and kindergarten boys were systematically rated lower by their teachers in academic and communication skills than same-age girls. Along the same lines, Berry, Shaywitz, and Shaywitz (1985) observed biases in the evaluation of students with attention deficit/hyperactivity disorder, favoring boys over girls.

Nevertheless, teacher biases must be more than an artifact of the methodology used (Flannery, Liederman, Daly, & Schultz, 2000). In an attempt to find out whether the evaluation method plays a significant role in that difference, Flynn and Rahbar (1994) reported that methods requiring subjective estimations lead to
higher proportions of special education referrals for boys than for girls, and that objective methods are more accurate than more subjective ones.

Biases in teacher ratings favoring girls over boys have been reported previously (Helwig, Anderson, & Tindal, 2001; Peterson & Bainbridge, 1999; Vogel, 1990; Wehmeyer & Schwartz, 2001). However, due to increased awareness, this effect may no longer be as prominent (Abidin & Robinson, 2002; Gillberg, 2003). For example, Goodman and Webb (2006) found no teacher gender bias in student referrals. Similarly, Helwig et al. noted that student gender was not a significant parameter in teachers’ evaluations of their students’ mathematics achievement. However, such bias in mathematics assessment does appear in later ages (e.g., secondary education) (Wimer, Ridenour, Thomas, & Place, 2001).

Thus, the findings from this line of research suggest that there may be biases when teachers rate boys compared to girls based on false expectations or other reasons. Shaywitz (1996) supported the finding that male vulnerability to reading disabilities is a result of referral and identification biases and has no genetic basis. Such biases may be due to the fact that boys tend to have more behavioral problems and attentional disorders than girls, which lead teachers to overrefer boys and clinicians to overdiagnose them with learning disabilities (Mamlin & Harris, 1998; Naiden, 1976; Share & Silva, 2003; Shaywitz et al., 1990; Ysseldyke et al., 1983).

Apart from students’ gender and behavior, which seem to be the main factors that influence teacher evaluations, Anderson (1997) concluded that students’ placement also plays a significant role through the mediating role of disruptive and inattentive behavior in the general classroom (Leinhardt, Seewald, & Zigmond 1982; Smart, Sanson, & Prior, 1996).

The present study investigated the presence of teacher biases with regard to identification of students with LD. More specifically, the study was designed to answer the following research questions in an attempt to further resolve current contradictory research findings:

1. Does teachers’ gender differentially affect their ratings with regard to identification of LD?
2. Are teachers’ age and experience, in addition to their gender, significant predictors of their student ratings?
3. Does children’s gender bias teacher ratings?

METHOD

Participants

The sample consisted of 246 Greek public school students, 33 with a diagnosis of LD and 213 typical students (125 boys, 121 girls). The students with LD were identified by a state multidisciplinary team of experts and referred for special education services. The definition of the disorder in Greece is identical to that of U.S. Department of Education currently used in the United States, identifying students as having LD when they demonstrate a significant discrepancy between potential and achievement. Nineteen of the students with LD were educated in inclusive general education classrooms. They were receiving special education services in addition to their education in typical (general education) settings.

The participating students were in grades 3 to 9 as follows: grade 3 = 33 (16 girls, 17 boys), grade 4 = 39 (17 girls, 22 boys), grade 5 = 39 (22 girls, 17 boys), grade 6 = 36 (18 girls, 18 boys), grade 7 = 29 (13 girls, 16 boys), grade 8 = 30 (15 girls, 15 boys), and grade 9 = 40 (20 girls, 20 boys). Two hundred and seventeen students spoke Greek as their native language; 29 were bilingual. Participating schools were selected by stratified sampling techniques in order to reflect the representative demographic areas of Greece. One boy and one girl were selected randomly from each class by excluding students who had a spoken language other than Greek as their native language or had not been attending a Greek school since first grade.

Student ratings were completed by 98 teachers during out-of-school hours, including 33 teachers of language arts, 63 elementary school teachers, and 2 from inclusive classrooms. With regard to gender distribution, there were 61 females and 31 males (data on gender were missing for 6 teachers). Teachers were distributed across grades as follows: grade 3 = 15, grade 4 = 19, grade 5 = 21, grade 6 = 18, grade 7 = 14, grade 8 = 17, and grade 9 = 19 (the total is larger than 100% because some teachers taught more than one grade). No significant effects were observed across distributions compared to the null model that all $\beta$ were equal (chi-square [6] = 3.496, p = .745). Teacher age ranged between 24 and 58 years, with a mean of 43.13 years. On average, teachers had 18.24 years of teaching experience. Sixty were teaching in elementary school units and 28 in secondary (data on setting were missing for 10 teachers). All teachers had fulfilled the requirements to teach in Greek public elementary or secondary schools.

Procedures

Teachers who were well acquainted with students’ work and achievement were asked to complete the Learning Disabilities Screening Scale for Teachers (Padeliadu & Sideridis, 2008). This scale is identical (in principle and concept) to the Learning Disabilities Diagnostic Inventory (LDDI) developed by Hammill (1995; see also Hammill & Bryant, 1998).

Ratings were collected during the last two months of the academic year to ensure that teachers were knowl-
edgeable of students’ behaviors and achievement levels. The latter was a prerequisite for teacher participation; that is, teachers had to feel confident that they could properly rate students’ performance and achievement. Teachers were not introduced to the concept of LD; instead they were asked to rate each item in terms of the frequency of that behavior (1 = always to 9 = never). Each subscale contained 17 to 20 items; the total number of items was 116. If a teacher was not able to complete a subscale (e.g., mathematical achievement in secondary schools), the teacher who best knew a given student’s achievement in the corresponding field was asked to complete the ratings.

The screening inventory refers to grades 3 through 9 and is composed of six subscales: listening, speaking, reading, writing, reasoning, and mathematics. These subscales correspond to specific areas of deficiency in the definition of learning disabilities from the National Joint Committee on Learning Disabilities (see Hammill, 1990). Specific behaviors involve, for example, the inability to discriminate between phonemes, comprehend text, and so on. Teachers supplied their own demographic information on a section of a given student’s rating scale.

Data Analysis

Multilevel random coefficient modeling (MRCM) was employed to evaluate whether a student’s diagnosis could be predicted by teacher gender and other characteristics (Kreft & de Leeuw, 1998; Roberts, 2003; Shin, Espin, Deno, & McConnell, 2004) or student characteristics. For that purpose, we fit the logistic multilevel model with a binary dependent variable (Boomsma, 1986). Further, in order for the resultant logistic coefficient to be translated into percentage points, we employed the following formula (Bryk & Raudenbush, 1992; Raudenbush & Bryk, 2002):

\[
\text{Probability of an Outcome} = \frac{1}{1 + \exp(-\eta_{ij})}
\]

with \(\eta_{ij}\) being the logistic regression coefficient. The specifics are discussed separately for each model in the following.

RESULTS

Unconditional Model of LD Identification

The following unconditional model was applied to the data to assess the mean ratings of LD prevalence using a population-based model (Choi, 2001; Raudenbush & Bryk, 2002; Singer & Willet, 2003), as follows:

**Level-1 Model**

\[
\text{Probability}_{\text{Diagnosis}} = \phi_{ij}
\]

\[
\log\left(\frac{\phi_{ij}}{1 - \phi_{ij}}\right) = \beta_{0j}
\]

**Level-2 Model**

\[
\beta_{0j} = \beta_{00} + \mu_{0j}
\]

with \(\Phi\) (phi) being the prevalence parameter, transformed into log units and expressed as \(\beta_{0j}\). Then prevalence rates are a function of their intercept \(\beta_{00}\), their error \(\upsilon_{0j}\), and any other predictor variables at level 2 of the model. In this model the level-1 variance is the reciprocal of the Bernoulli variance \(\gamma_{1j} = (1 - \gamma_{1j})\).

Results indicated that 9.3% of the students were identified as having LD. However, we bootstrapped the population mean (Chernick, 2007; Efron, 1982) to ensure that there were not substantial biases due to participant selection (Diaconis & Efron, 1983).

Bootstrapping the prevalence of LD in the population.

Using a population-based model, the prevalence of LD in the Greek population was estimated to be 9.3%. However, to ensure that the sample’s estimates of prevalence were unbiased, we bootstrapped the prevalence parameter using a nonparametric bootstrap and applying 1,000 replications. Results indicated that the sample’s mean prevalence rates were 9.302% compared to 9.286% in the bootstrap distribution (i.e., the sampling distribution based on 1,000 replications) showing a bias of -0.0001, which is negligible.

Figure 1 shows the sampling distribution of the prevalence data, indicating that the mean rate of identification was indeed an unbiased estimate of the population effects. This analysis further substantiates subsequent findings related to the biasing effects of teacher gender on identification of learning disabilities. Using confidence intervals derived from the bootstrap distribution, the prevalence rates at 95% ranged between 5.58 and 13.02.

Bias Due to Teacher Gender

In order to test for teachers’ biases due to their gender, the following multilevel logistic model was fit to the data using the Bernoulli function in HLM 6.1 (Raudenbush & Bryk, 2002):

**Level-1 Model**

\[
\text{Probability}_{\text{Diagnosis}} = \phi_{ij}
\]

\[
\log\left(\frac{\phi_{ij}}{1 - \phi_{ij}}\right) = \beta_{0j} + \beta_{1j}(\text{Teacher Gender}) + \mu_{0j}
\]

**Level-2 Model**

Given that teacher gender was coded as 0 = females and 1 = males, the intercept reflects the estimates for males and the slope reflects those of female teachers. Results indicated that the log estimate for females was -1.98 units and that of males -1.06 units. The former parameter was significantly different from zero, whereas the latter approached significance but did not exceed it \((p = 0.90)\). This suggests that the prevalence rates of
male teachers identifying children as having LD were significantly different from zero. When transforming these logits into percentages using Equation 1, the probability of classifying students as having LD was 12.13% for female teachers and 25.7% for male teachers. Given that population estimates were slightly below 10% (i.e., 9.3%), it was concluded that male teachers produced inflated judgments regarding identification of students having LD compared to female teachers. The bias introduced by teacher gender was substantial, and for that reason a series of alternative hypotheses were investigated to ensure that this finding is real and not an artifact of other teacher characteristics (i.e., teaching experience or age).

**Bias Due to Student Gender**

A number of studies have shown that student gender influences teacher judgments (Rivard et al., 2007). In order to evaluate the potential differential effects of students’ gender, the following model was fit to the data:

**Level-1 Model**

\[
\text{Probability}_{\text{Diagnosis}} = \beta_0 + \beta_1 (\text{Gender of Child}) + r_i
\]

**Level-2 Model**

\[
\beta_0 = \gamma_0 + u_{0i}
\]

\[
\beta_{1i} = \gamma_{10} + \gamma_{11} (\text{Gender of Teacher})
\]

The coefficients of interest here are \(\beta_{1i}\) and \(\gamma_{11}\). The term \(\beta_{1i}\) evaluates whether students’ gender affects the probability of identification. In other words, whether more girls than boys are identified as having LD. The term \(\gamma_{11}\) evaluates whether teacher gender affects the relationship between students’ gender and diagnosis. In other words, the model tests whether teachers’ gender moderates the relationship between students’ gender...
Results indicated that both coefficients were non-significant. Specifically, \( \beta_{11} = -0.019, p = 0.367 \), suggesting that the probability of having LD was independent of a child’s gender. Also, the term \( \beta_{11} = 0.016, p = 0.448 \) was nonsignificant, suggesting that the relationship between a child’s gender and the probability of an LD diagnosis was not affected by teacher gender (i.e., by the teacher who provided the ratings).

**Contribution of Teacher Age and Experience**

In order to test whether teacher characteristics such of age and years of experience working with students with special needs affected their judgments, the following multilevel logistic model was fit to the data:

**Level-1 Model**

\[
\text{Probability}_{\text{diagnosis}} = \phi_{ij}
\]

\[
\log(\phi_{ij} / (1 - \phi_{ij})) = \beta_{0j}
\]

**Level-2 Model**

\[
\beta_{0j} = \beta_{00} + \beta_{01} \text{(Teacher’s Age)} + \beta_{02} \text{(Teacher’s Experience) } u_{0j}
\]

with all \( \beta_i \) representing partial regression coefficients associated with teachers’ age and experience, controlling for all other variables in the model. Age and teaching experience were entered as a grand mean centered, so the results are reflected as if variables had undergone a z-score transformation. Results, as shown in Table 1, indicated that neither of these demographic variables was a significant predictor of prevalence rates.

**DISCUSSION**

This study investigated the presence of teacher biases with regard to identification of students with learning disabilities. Factors related to teachers’ gender, age, and experience, along with students’ gender, were investigated. Results indicated that teachers’ gender significantly biases their identification of learning disabilities, with females being more lenient than males. The bias was in the magnitude of 2:1, with male teachers identifying two students as having LD for every one student identified by female teachers.

This finding may be interpreted to mean that female teachers initiate more interactions with their students and especially pay more attention to boys (Einarsson & Granstroem, 2002). With regard to the more “positive” evaluations by female teachers, Noddings (1984) and Gilligan (1982) noted that woman teachers are more caring (Einarsson & Granstroem), supportive, and expressive (Hopf & Hatzchistou, 1999) than men. Additionally, Einarsson and Granstroem found that male teachers pay less attention to male students as they grow older, a fact that may offer a logical explanation of the results of this study. Thus, overrepresentation of LD may partly be due to the gender of the teacher (if we consider that the ratings of male teachers were enhanced compared to the baseline model).

This finding agrees with previous ratings showing substantial biases based on teacher characteristics such as psychopathology (Lovejoy, 1996) but not other char-

---

**Table 1**

**Multilevel Random Coefficient Model Predicting Prevalence of Identification Rates Due to Teacher Age and Years of Experience**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient</th>
<th>S.E.</th>
<th>t-ratio</th>
<th>p</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept ( \beta_0 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept ( \beta_{00} )</td>
<td>-2.118</td>
<td>2.466</td>
<td>-0.859</td>
<td>0.394</td>
<td>68</td>
</tr>
<tr>
<td>Level-2 Model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher’s Age ( \beta_{02} )</td>
<td>-0.016</td>
<td>0.067</td>
<td>-0.240</td>
<td>0.811</td>
<td>68</td>
</tr>
<tr>
<td>Teacher’s Experience ( \beta_{03} )</td>
<td>0.030</td>
<td>0.040</td>
<td>0.740</td>
<td>0.862</td>
<td>68</td>
</tr>
</tbody>
</table>

*Note. Coefficients reflect fixed effects of the model. S.E. = Standard error of measurement.*

---

*Volume 31, Fall 2008  205*
acteristics such as teaching specialty (general vs. special education) (Liljequist & Renk, 2007). However, the present findings disagree with those of Rescorla et al. (2007) with a different population (i.e., students with emotional problems) in that students referred for emotional problems did not lead to substantially different teacher ratings.

A second important finding was that teachers’ age and experience were not significant predictors of their identification ratings. This finding is troublesome and contradicts the literature, which points to these factors being important teacher resources (Liljequist & Renk, 2007).

A third important finding relates to the fact that students’ gender was not associated with differential teacher ratings. In addition, these ratings were not related to teacher gender. This agrees with past research in that the gender of the child was not a significant determinant of teachers’ ratings of a disability (Rivera et al., 2007). This finding is extremely important given prior research confirming that teacher biases create mindsets and can be devastating for the development of prejudice in specific groups or populations (Rosenthal & Jacobson, 1968).

Limitations

Despite important findings, the present study is limited by a number of factors. First, although all students had a diagnosis of LD by an approved authority, it is well known that teams employ variable criteria (e.g., use different standardized assessments or place more emphasis on Factor A than Factor B). For example, teams with experts on speech and language pathology may rely more heavily on such measures than teams that apply a discrepancy (Stanovich, 2005) or an RTI model (Fuchs & Deshler, 2007; Haager, 2007; Vaughn & Fuchs 2003). Thus, although all students had a diagnosis of LD, we cannot verify the validity of these ratings beyond the fact that students demonstrated significant sub-average achievement.

A second limitation pertains to the fact that only teachers’ age, gender, and experience were tested. Thus, it is likely that other demographic variables may have been responsible for the observed gender effect. Inclusion of such variables may aid our understanding of the existence of biases by teachers. For example, Pugach and Johnson (1995) and Tournaki and Podell (2005) reported that low numbers of referrals by teachers were predicted by their efficacy ratings; nevertheless, the opposite finding was reported by Soodak and Podell (1993, 1996).

Furthermore, as two thoughtful reviewers of this article suggested, the ratings by teachers may reflect not just their observations but other sources of influence as well. Unfortunately, this is always the case with ratings that may contain both a subjective and an objective part. Thus, ratings likely reflect a knowledge base that is due to various sources, including observations. If other sources of influence bias teachers’ ratings, they should be systematically examined in a methodological study (beyond the scope of the present study). Nevertheless, teacher ratings represent a reality in our current use of evaluative criteria in learning disabilities. If other sources influence these ratings (besides those examined in the present study), we hope that they are randomly distributed across teachers, thus adding to the unsystematic source of error (i.e., random error).

Implications for Practice

The fact that the present study found biases in teachers is troublesome; especially, given the mandate by the No Child Left Behind Act (2001) for highly qualified teachers (King-Sears, 2005). Teachers are increasingly relied on to rate students’ achievement and behavior. As a result, if their ratings are lacking in validity, it can have serious implications for students’ placement (Fierros & Bloomberg, 2005) and subsequent well-being. Although the present findings using Greek teachers and students cannot be directly generalized to U.S. teachers, it is plausible that such an effect may be present given the resemblance in the two educational systems and teachers’ education (the Greek system has borrowed the main components of the U.S. educational system). Nevertheless, U.S. studies are needed to ensure that such an effect is present.

Teacher awareness and training may be avenues for improving the accuracy of their ratings (Moore, 2008; Rivard et al., 2007). Teacher efficacy has proven to be one of the most important factors contributing to the validity of teacher judgments (Einarsson & Granstroem, 2002; Shinn, Tindal, & Spira, 1987). That is, teachers with low personal efficacy are more likely to mistakenly refer students for special education services (Einarsson & Granstroem). Training and practical guidance may enhance teachers’ personal efficacy and, in turn, influence the validity of their estimates. The U.S. Department of Education’s emphasis on enhancing teachers’ content knowledge is important in this connection (Boe, Shin, & Cook, 2007). Other suggestions for improving the validity of teacher ratings involve the use of brief, rather than lengthy, assessments (Henderson & Sugden, 1992) and instruments with documented positive psychometric properties (Junaid, Harris, Fulmer, & Carswell, 2000).

As Rivera et al. (2007) pointed out, “Teachers have a critical role to play in the identification and management of children” (p. 645) with and without disabilities. Future studies could enrich our understanding of
the factors that contribute to valid teacher ratings or enhance the efficacy of teachers to be successful in their multidisciplinary role (Dembo & Gibson, 1985; Gibson & Dembo, 1984; Merrett, & Wheldall, 1993).

REFERENCES


Please address correspondence regarding this article to: Georgios D. Sideridis, University of Crete, Department of Psychology, Rethimno, 74100 Crete, Greece; e-mail: sideridis@psy.soc.uoc.gr